IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR U.S. LETTERS PATENT

Title:

DISTRIBUTED AUDIO COLLABORATION METHOD AND APPARATUS

Inventor:

Steven D. Curtin

Thomas J. D'Amico
DICKSTEIN SHAPIRO MORIN &
    OSHINSKY LLP
2101 L Street NW
Washington, DC 20037-1526
(202) 828-2232

TITLE OF INVENTION:

DISTRIBUTED AUDIO COLLABORATION METHOD AND APPARATUS

5   FIELD OF THE INVENTION

The present invention is directed to a method and apparatus which allows audio signals from different sources to be combined and replayed back at the origination source of each audio signal.

10   BACKGROUND OF THE INVENTION

Many times musicians and/or vocalists desire to collaborate to form a musical work without the need for assembling all musicians/vocalists in a recording studio. Thus, the audio signal from different musical instruments, vocalists and/or other audio sources can be recorded individually at a location,

15   then later mixed together to form a composite musical work, and then sent back to the musicians/vocalists. However, such activity cannot be performed in real time, and multiple musicians and/or vocalists wishing to collaborate in real rime from multiple locations cannot do so.

There are, however, conference call systems which allow multiple

20   users wishing to make a conference call to do so without the use of an operator or a bridge number. However, such conference call devices do not allow multiple musicians/vocalists to collaborate in real time from multiple locations to form a composite work which each musician/vocalist can hear at the same time. Musicians/vocalists attempting to collaborate by a conference call are also

25   severely limited by the audio constraints of the telephone system, which typically

1230915 v1; QDS301!.DOC

does not pass any audio signals above 3000 - 4000 hertz, thereby providing a significant limitation on audio quality.

In an effort to more quickly form composite musical works, program servers have been developed which interface with the Internet and allow multiple
5   musicians at different locations to use the Internet to send a MIDI audio streams over the Internet to the server, which mixes the audio sources using an MIDI merge function and feeds the merged MIDI signal back to participating musicians. However, this system, too, does not operate in real time, and cannot provide feedback to the musicians/vocalists while they are playing their
10   instruments and/or singing their vocal part.

What is needed is a method and apparatus which can allow multiple musicians/vocalists at various locations to easily collaborate on a musical work and which provides near real time feedback of the collaborative work at the multiple locations at which the musicians/vocalists are located. What is also
15   needed is a simplified high fidelity conference call system which does not require operator interaction to establish a conference connection.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides a method and apparatus
20   operating over a communications network, e.g., the Internet, which permits conference calls to be easily made by the streaming transmission of compressed audio signals from each participant client location to a server location where the individual audio signals are decompressed, combined into a composite audio signal, compressed and broadcast back to each of the client locations for
25   decompression and play back. The composite audio signal can be a

concatenation of the individual audio signals or a mix of the individual audio signals. The method and apparatus operates in near real time.

5   In another aspect, the present invention provides a distributed audio collaboration method and apparatus operated over a communications network, e.g., the Internet, in which each of a plurality of musicians/vocalists can transmit in streaming fashion their respective audio contributions to a collaborative musical work in compressed form from a client computer to a common server computer, which decompresses the audio contributions and combines the individual contributions into a composite audio signal which is in turn

10   compressed and broadcast in streaming fashion by the server over the communications network back to each of the musicians/vocalists client computers which can decompress the composite audio signal for play back. The composite audio signal may be in the form of a concatenation of the individual audio contributions provided by each of the musicians/vocalists without any

15   mixing of the contributions by the server, or the server can perform an actual mix of the audio contributions provided by each of the musicians/vocalists and provide the composite audio signal as a composite mix back to each of the musicians/vocalists. If the composite signal is a concatenation of the individual works, each musician/vocalist may at his own client computer location mix the

20   received concatenated signals as desired. The audio collaboration method and apparatus operate in near real time.

These and other features and advantages of the invention will be more clearly understood from the following detailed description which is provided in connection with the accompanying drawings.

25

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows a typical network, e.g., the Internet, which may be used to implement the invention;

FIGURE 2 is a block diagram of a server computer illustrated in
5   Figure 1 which is used to concatenate audio signals from a plurality of client computer sources;

FIGURE 3 is a block diagram of a server computer illustrated in Figure 1 used to mix a plurality of audio signals from a plurality of client computer sources;

10   FIGURE 4 is an illustration of a ping technique used between client and server computers illustrated in Figure 1 for determining transmission delays;

FIGURE 5 is a block diagram of a client computer which receives a concatenated audio transmission from a server computer; and

FIGURE 6 illustrates a client computer which receives a previously
15   mixed composite audio transmission from a server computer.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides a network based method and apparatus for combining audio signals from different sources at respective client computers
20   into a composite audio signal at a central server computer, and then providing the composite audio signal back to the individual client computers. The composite audio signal may be created by merely concatenating individual audio signals received from the client computers, and broadcasting the concatenated signals back to the client computers for mixing and replay, or the server
25   computer may mix the received audio signals to form a composite mixed audio

signal which is broadcast back to the client computers for replay. The method and apparatus may be used for conference calling, or for musician/vocalist collaboration in forming a collaborative musical work.

5   For simplicity, the method and apparatus of the invention will be described below in the context of forming a collaborative musical work in which the originating audio signals are from musical instruments or vocalists. However, the same illustrated and described method and apparatus may be used for conference calls with the originating audio sources being voice signals.

10   Figure 1 illustrates a networked environment in which the invention is used. Each of the musicians and/or vocalists are located at an audio node client computer 11 which is coupled through a communication link to a server computer 13. The communication link can be any digital transmission path including, for example, a wide area network or other network, but is preferably the Internet, indicated in Figure 1 by numeral 15.

15   Each of the audio node client computers 11 is provided with streaming software which receives an audio signal source at the client computer 11 and converts it into an encoded (compressed) stream of digital packets which are sent over Internet 15 to server computer 13. The encoded audio packets are sent from the client computers 11 in real time, and the server computer 13 receives the encoded audio signals from each of the client computers 11, decodes (decompresses) them, and forms therefrom a composite audio signal which contains the audio signal streams from each of the client computers 11. The server computer encodes (compresses) the composite audio signal and transmits the encoded composite signal back to each of the client computers 11.

25   As a consequence, a musician/vocalist at a particular client computer may play his/her instrument or sing his/her part, and at the same time obtain

near real time feedback from the server computer 13 of a composite audio signal, which includes his/her contribution, as well as the contributions of other musicians/vocalists at other client computers 11. Near real time refers to real time minus delays associated with the network transmission and computer

5   processing functions.

The composite audio signal which is sent back from the server computer 13 either can be in the form of a concatenation of the audio signals from each of the client computers 11, that is encoded packet data which merely successively attaches the audio signals from each of the client computers 11

10   without mixing them, or a mixed audio signal formed by server computer 13 mixing the audio signals from the client computers 11 and providing a composite mix signal as packet data to each of the client computers 11.

Figure 2 illustrates the structure and operations performed at server computer 13 when providing a concatenation of all received audio signal packets.

15   As illustrated in Figure 2, a plurality of users identified as User 1 . . . User 4, provide encoded (compressed) audio signals at each of their client computers 11. These encoded audio signals are received by the server computer 13 and are decoded (decompressed) by decoder 23 which is associated with each of the received audio signals, into respective audio signal samples. The samples are in

20   turn fed to a respective linear delay line 25 which compensates for individual transmission delays of the respective audio signals, and the delayed samples are then sent to software within server computer 13 which concatenates, i.e., serially combines, all of the samples together in a concatenation module 26. That is, a long stream of sequential samples from each of the users is formed. This

25   concatenation signal is then encoded (compressed) at the server computer 13 by encoder 29, and is then in turn rebroadcast over the Internet 15 back as a single signal in common to all of the client computers 11.

The manner in which the client computers 11 handle a composite audio signal formed by the concatenation of all user audio signals will be described below in connection with Figure 5.

Figure 3 illustrates the structure and operation of server computer 13 when it performs a mixing operation for the received audio signals from the users. The encoded audio signals from the individual client computers 11 are received at server computer 13 and each is decoded (decompressed) by decoder 23. Again, as in Figure 2, the server computer 3 provides a linear delay line 25 for each decoded audio signal which compensates for individual delays in transmission of the received audio signals and provides the audio samples to a mixer module 31.

The mixer module 31 linearly sums the delayed outputs of the decoders 23, and provides a single mixed composite signal in the form of samples to an encoder 29, which encodes (compresses) those samples and then broadcasts them back to all client computers 11.

Figures 2 and 3 illustrate linear delay lines 25 which are provided in each of the audio signal receiving paths at server computer 13. The linear delay lines 25 are individually adjustable to time align the received audio signals from the various users in accordance with the transmission delays associated with each user. This way the audio signals from the various users are synchronized. In order to determine the appropriate delay for the received signals from a given user, a ping technique illustrated in Figure 4 is employed.

In the ping technique shown in Figure 4, the server computer 13 sends a ping message to a client computer 11. Upon receipt of the ping message, the client computer 11 time stamps the message and resends it in a ping message back to the server computer 13. The server computer 13, upon

receiving the time stamp message from the client computer 11, compares the time stamp with the time at which the message is actually received, the time difference representing the transmission time for a message to travel from client computer 11 to server computer 13. This pinging is done for each of the client

5   computers 11 illustrated in Figure 1, and the server 13 accordingly adjusts each of the linear delay lines 25 for the received audio signals to ensure that all are properly time aligned by eliminating effects of transmission delays in the time of receipt of the respective audio signals.

Figure 5 illustrates the structure and operations of the client

10   computers 11 when the server computer 13 concatenates the individual received audio signals. A source 41 of electronic and/or acoustic musical instruments and/or vocals provides digital samples of an input audio analog signal to an audio encoder 43. Audio encoder 43 compresses the audio signal using any one of many conventional audio signal compression techniques , e.g., MP3, PAC

15   (perceptual audio coder), Dolby AC-3 , MPEG-4, etc. The encoded audio signal is then sent over the Internet 15 to the server computer 13. The client computer 11 also receives from the Internet 15 the encoded audio composite music signal which, in the case of the Figure 5 implementation, represents the concatenation of the individual audio sources by the server computer 13.

20   The audio signals received at client computer 11 are then decoded (decompressed) at decoder 45, and the decoded samples are then sent to a user controlled linear mixing device 47, and the audio samples from the user controlled linear mixing device are then fed to a monitoring device 49 such as an audio reproduction circuit where the digital audio samples are converted to an

25   analog audio signal for replay over one or more audio channels and associated speakers. It should be noted that the user has his/her own control over the

linear mixing in mixing module 47, and can appropriately adjust the mixing conditions to produce a desired audio output at the monitoring device 49.

Figure 6 illustrates the client computer 11 when configured to operate with a server computer 13 which mixes the audio signals received from

5   the client computers 11. Once again, a source of electronic and acoustic musical instruments and/or vocals is provided at source 41 which provides digital samples of an analog audio signal to an audio encoder 43 which transmits the encoded audio signal over the Internet 15 to computer server 13.

Figure 6 also illustrates the reception at client computer 11 of the

10   Internet 15 transmission from the server 13. In the Figure 6 arrangement, the server computer 13 has already mixed the audio signals from the individual client computers 11 so that a mixed signal is received at the client computer 11. This received mixed signal is then decoded by decoder 45, and decoded samples are fed to a musician monitoring device 49 which includes an audio circuit and one

15   or more speakers, as described above with reference to Figure 5.

As noted above, the invention allows a musician/vocalist to collaborate with other musicians/vocalists at diverse locations in near real time to create a collaborative musical work which all musicians/vocalists can receive and hear at nearly the same time they are making their contribution to the

20   collaborative work. Since the audio signals can be sampled and transmitted with high fidelity, the resulting composite work as replayed at the musician sites is likewise of high fidelity.

For lower bandwidth applications and in order to allow lower bit rate audio encoding and decoding, and to provide a uniform standard for

25   communication of musical gestures, all music information can be provided in MIDI format or other structured audio formats, such as C-sound or MPEG-4

SAOL at both the client computers 11 and server computer 13 rather than being encoded/decoded. In this case, the server computer 13 when mixing the audio signals can use a conventional MIDI merge function. Moreover, although the invention has been described as using exemplary MP3, PAC, Dolby AC-3 and

5   MPEG-4 encoding/decoding at the client computer 11 and server computer 13 locations, other conventional and available audio encoding/decoding techniques can be used as well.

While the invention has been described and illustrated with respect to exemplary embodiments, it should be understood that various modifications,

10   substitutions, deletions and additions can occur without departing from the spirit and scope of the invention. Accordingly, the invention is not to be considered as limited by the foregoing description, but is only limited by the scope of the appended claims.